nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-018-0200-0

Bipartite interactions, antibiotic production and biosynthetic potential of the *Arabidopsis* leaf microbiome

Eric J. N. Helfrich[1,2], Christine M. Vogel[1,2], Reiko Ueoka[1], Martin Schäfer[1], Florian Ryffel[1], Daniel B. Müller[1], Silke Probst[1], Markus Kreuzer[1], Jörn Piel[1*] and Julia A. Vorholt[1*]

Plants are colonized by phylogenetically diverse microorganisms that affect plant growth and health. Representative genome-sequenced culture collections of bacterial isolates from model plants, including *Arabidopsis thaliana*, have recently been established. These resources provide opportunities for systematic interaction screens combined with genome mining to discover uncharacterized natural products. Here, we report on the biosynthetic potential of 224 strains isolated from the *A. thaliana* phyllosphere. Genome mining identified more than 1,000 predicted natural product biosynthetic gene clusters (BGCs), hundreds of which are unknown compared to the MIBiG database of characterized BGCs. For functional validation, we used a high-throughput screening approach to monitor over 50,000 binary strain combinations. We observed 725 inhibitory interactions, with 26 strains contributing to the majority of these. A combination of imaging mass spectrometry and bioactivity-guided fractionation of the most potent inhibitor, the BGC-rich *Brevibacillus* sp. Leaf182, revealed three distinct natural product scaffolds that contribute to the observed antibiotic activity. Moreover, a genome mining-based strategy led to the isolation of a *trans*-acyltransferase polyketide synthase-derived antibiotic, macrobrevin, which displays an unprecedented natural product structure. Our findings demonstrate that the phyllosphere is a valuable environment for the identification of antibiotics and natural products with unusual scaffolds.

Natural products are a remarkably diverse group of metabolites with relatively simple to highly complex structures. Most are classified as specialized metabolites that, unlike primary metabolites, are usually not directly involved in growth, reproduction or survival[1]. These compounds are typically restricted to a few producers and serve to interact with the local environment. The structural diversity of natural products is reflected in a wealth of different bioactivities, many of which have been exploited as drugs in human and veterinary medicine. The majority of approved small-molecule drugs, and especially antibiotics, are either natural products, natural product-derived or inspired by natural products[2]. The ecological role of specialized microbial metabolites is in most cases only poorly understood. Some play a role in microbial warfare[3], whereas others function as developmental signals[4,5], act as signalling metabolites in host–microbe or microbe–microbe interactions[3,6,7] or serve as chelators to acquire metals from the local environment[8]. Recent studies have shown that so far neglected ecological niches are treasure troves for the discovery of chemical novelty[9–11]. Thus, the analysis of microorganisms in underexplored habitats represents a promising path to identify natural products and to investigate their roles in situ. Metagenomics-based studies of such environments have resulted in the successful isolation of natural products[12–14]. Yet, this approach does not provide direct access to the producing organisms.

Diverse culture collections of bacteria are a valuable resource to examine the potential of natural product formation both via bioactivity assays and genome-mining[15]. Recently, the combination of culture-independent microbiota profiling with large-scale bacterial isolation efforts resulted in the successful cultivation of exemplars of the majority of species that are reproducibly associated with leaves and roots of natural *Arabidopsis thaliana* populations[16]. These strain collections, and their respective draft genome sequences, are now available for systematic experimental and in silico analyses.

To assess the biosynthetic potential of this resource, we conducted a large-scale interaction screen, together with bioinformatic and chemical analyses, focusing on the leaf microbiota. From a biological perspective, the limited availability of nutrients in this habitat[17,18] might have selected for highly competitive strains engaged in chemical warfare. We present a comprehensive interaction screen of more than 200 strains (the *At*-LSPHERE collection)[16] tested against each other and selected pathogens. Our results reveal numerous inhibitory interactions and show that some strains inhibit a large phylogenetic diversity of strains. Genome analysis identified about 1,000 BGCs covering a large variety of compound classes. *Brevibacillus* sp. Leaf182, a 'talented producer' identified by the interaction screen as well as in silico predictions, was selected for in-depth chemical analyses in order to characterize metabolites responsible for the observed high inhibitory activities against ~50% of all tested strains.

Results

Binary interaction network of phyllosphere bacteria. To survey the network of inhibitory interactions within a collection of more than 200 bacterial leaf isolates, we screened the *At*-LSPHERE strain collection[16] for inhibitions in binary combinations. Each strain was tested for potential growth interference against all members of the strain collection (Fig. 1a). In total, we monitored more than 50,000 pairings of the 224 strains belonging to the main bacterial

[1]Institute of Microbiology, ETH Zurich, Zurich, Switzerland. [2]These authors contributed equally: Eric J. N. Helfrich, Christine M. Vogel.
*e-mail: jpiel@ethz.ch; jvorholt@ethz.ch

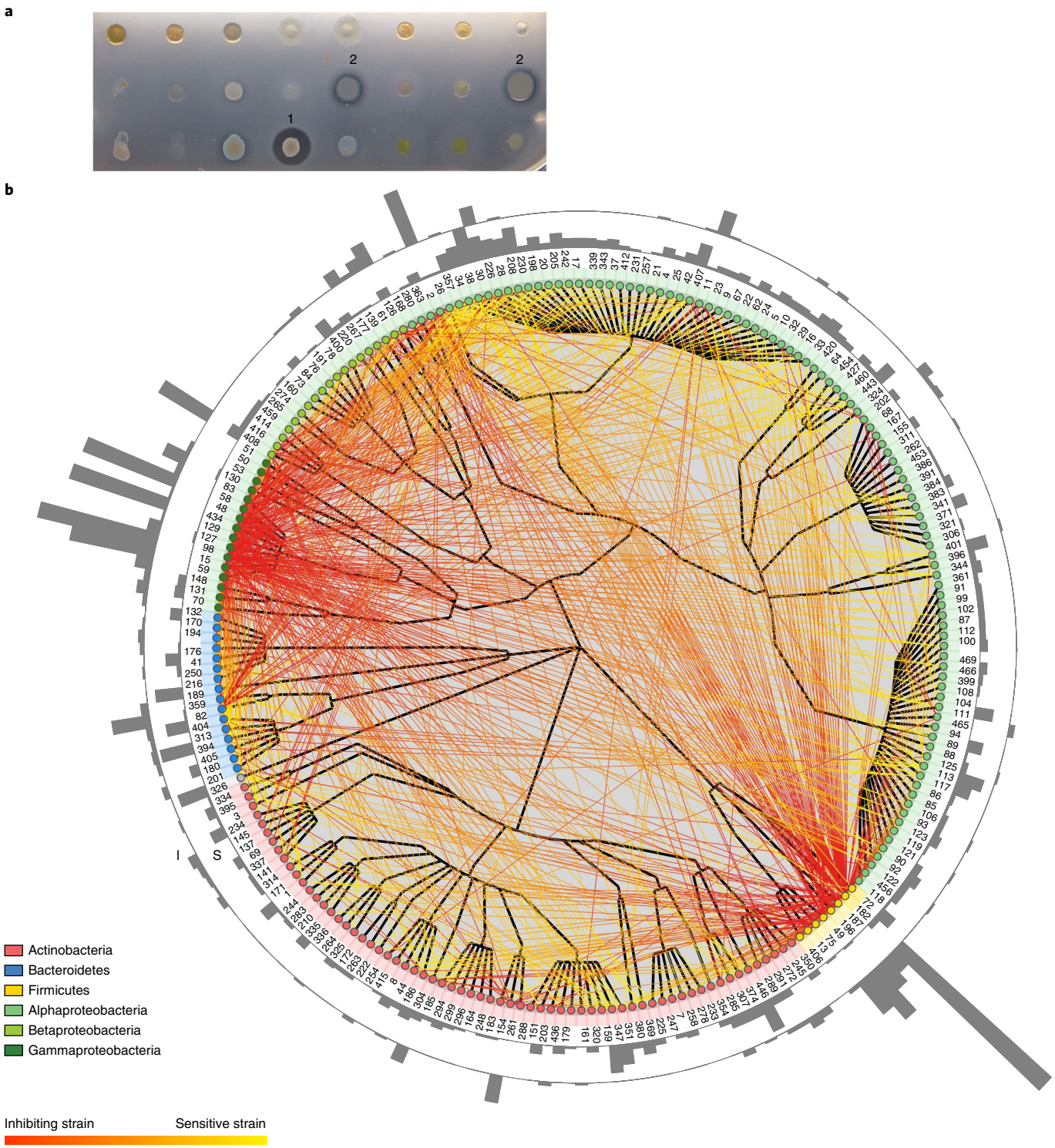

Fig. 1 | *Bacillales* and *Pseudomonadales* dominate the binary interaction network. a, Exemplary phenotypes of interaction classifications. Inhibitory halos are labelled with 1 for strong and 2 for weak inhibitions, based on halo size. **b**, Taxonomic representation of all isolates ($n = 224$) with observed directional interactions. Numbers in the inner ring indicate strain numbers and each node is colour coded according to phylum or class annotation, whereas the annotation background colour represents phylum assignment. Directional inhibitory interactions (725 of 50,176 pair-wise combinations tested, 1–2 biological replicates) are indicated from red to yellow. The number of sensitivities per strain is summarized in the innermost circle (S), and the number of inhibitions caused in the outermost ring (I). For a list of all interactions see Supplementary Table 1.

phyla inhabiting the *A. thaliana* phyllosphere. Notably, 725 inhibitory binary interactions (1.4% of all possible pairings) were observed (Supplementary Table 1). Overall, 196 strains (88%) engaged in inhibitory binary interactions (Fig. 1b, Supplementary Fig. 1). Of these, 160 strains (71%) were sensitive against other phyllosphere isolates and 79 strains (35%) inhibited 1 or multiple strains, with 26 strains showing inhibition of 5 or more strains (Supplementary Table 2); notably, many of these inhibited isolates belonging to

Table 1 | The ten *At*-LSPHERE strains showing most inhibitory interactions in the binary interaction screen of 224 strains against 224 strains (1–2 biological replicates)

Rank	Strain	Phylum	Genus	Number of inhibitions		
				Total	Strong	Weak
1	Leaf182	Firmicutes	*Brevibacillus*	111	32	79
2	Leaf98	Proteobacteria	*Pseudomonas*	67	17	50
3	Leaf434	Proteobacteria	*Pseudomonas*	56	6	50
4	Leaf58	Proteobacteria	*Pseudomonas*	54	4	50
5	Leaf15	Proteobacteria	*Pseudomonas*	47	10	37
6	Leaf49	Firmicutes	*Bacillus*	32	9	23
7	Leaf2	Proteobacteria	*Novosphingobium*	31	5	26
8	Leaf408	Proteobacteria	*Methylophilus*	31	11	20
9	Leaf75	Firmicutes	*Bacillus*	22	9	13
10	Leaf82	Bacteroidetes	*Flavobacterium*	21	9	12

different phylogenetic clades (Supplementary Table 3). Most isolates, however, inhibited only one or two other strains, suggesting specific interactions. Cluster analysis of all bipartite interactions indicated that interactions occur across all phylogenetic clades (Supplementary Fig. 2), with the 10 most active strains (>20 inhibitions each) forming the central nodes of the interaction network (Fig. 1, Supplementary Fig. 2, Table 1).

Inhibitory interactions might be context dependent and rely on sufficient growth, specific stimuli or other environmental conditions. We thus selected ten strains belonging to different phylogenetic clades and tested whether inhibitions of these by all *At*-LSPHERE strains are robust to alternative growth media. We used six media differing in their complexity, with some mimicking the nutrient sources available on leaf surfaces (see Methods). Although not all interactions could be scored due to poor growth, overall, the observed inhibitions were rather congruent among all media tested (Supplementary Fig. 3, Supplementary Tables 4 and 5).

The entirety of all bacterial interactions revealed that the majority of observed inhibitions were due to the activity of two bacterial orders, *Bacillales* and *Pseudomonadales* (Fig. 1b, Supplementary Figs. 1 and 2). These two taxa make up 8% of the strain collection but confer over 60% of observed inhibitions. *Bacillales* isolates frequently inhibited *Sphingomonadales* (22% of possible inhibitions), *Caulobacterales* (17%) and *Actinomycetales* (12%), whereas strains of the order *Pseudomonadales* often inhibited *Methylophilales* (28% of possible inhibitions), *Flavobacteriales* (19%), *Sphingomonadales* (18%), *Xanthomonadales* (17%) and *Rhizobiales* (14%) (Supplementary Table 6). *Sphingomonadales* were more often inhibited by *Bacillales* and by *Pseudomonadales* than expected based on the overall observed inhibitions (Supplementary Table 6). A closer inspection of the top ten inhibiting strains (Table 1) revealed that the two Firmicutes, *Brevibacillus* sp. Leaf182 and *Bacillus* sp. Leaf49, as well as *Flavobacterium* sp. Leaf82, inhibited more members of the genus *Sphingomonas* (Alphaproteobacteria) than expected (one-sided Fisher's exact test, Bonferroni-adjusted $P=0.003$, $P=0.01$ and $P=0.0007$, respectively), whereas *Pseudomonas* sp. Leaf15 was active against many members of the genus *Methylobacterium* (Supplementary Table 7, Supplementary Fig. 4).

Overall, only a minority of interactions were observed between members of the same family or genus. One exception was members of the genus *Aeromicrobium*, which inhibited the *Nocardioides* and *Marmoricola* within the *Nocardioidaceae* (Supplementary Table 6). Furthermore, 14 inhibitory activities between strains belonging to the same genus (0.4% of 3,264 possible combinations) were observed, significantly less than expected by chance (two-sided Fisher's exact test, $P=8.8\times10^{-9}$). Overall, this result indicates that leaf isolates tend to inhibit distinct phylogenetic groups rather than closely related strains.

To assess the inhibitory potential of phyllosphere isolates against well-studied bacterial phytopathogens, we also tested *Pseudomonas syringae* pv. *tomato* DC3000, *P. syringae* pv. *syringae* B728a, *Ralstonia solanacearum* AW1, *Agrobacterium tumefaciens* C58 and *Xanthomonas campestris* pv. *campestris* LMG 568. Several of the top inhibitory strains also showed activity against these model pathogens (Supplementary Table 8). *Pseudomonas* sp. Leaf58 showed inhibitory activity against all tested pathogens except *R. solanacearum* AW1, whereas *Pseudomonas* sp. Leaf434 inhibited all but *P. syringae* DC3000 and *R. solanacearum* AW1. In consequence, strains of the collection isolated from healthy plants have potential to help in defence against pathogens via direct inhibition.

Identification of putative natural product BGCs. The identification of a large inhibitory interaction network prompted us to investigate the full biosynthetic potential of each strain in the *At*-LSPHERE[16] collection. We thus mined the genome-sequenced strains for BGCs putatively involved in the production of natural products. The antibiotics & Secondary Metabolite Analysis Shell (antiSMASH) 4.0 toolkit[19] was used for the prediction and annotation of natural product BGCs. In total, 1,053 BGCs were identified and separated into 8 distinct natural product (sub)classes (Fig. 2, Supplementary Tables 9–11). In order to identify BGCs that might encode the biosynthesis of uncharacterized natural product scaffolds, antiSMASH results were subsequently compared against the MIBiG database[20] of characterized BGCs using the Biosynthetic Genes Similarity Clustering and Prospecting Engine (BiG-SCAPE) (https://git.wageningenur.nl/medema-group/BiG-SCAPE) (Supplementary Table 12), a recently established analysis platform to compare BGCs based on distance metrices. BiG-SCAPE analysis revealed a large set of 766 BGCs (73%) that putatively encode biosynthetic novelty (Fig. 3, Supplementary Fig. 5, Supplementary Table 13), as indicated by orphan BGCs that are not connected to characterized entries in the MIBiG database. In total, 97 clusters that did not contain any MIBiG BGCs and 176 singletons were found.

Among the most important natural product classes that contain specialized metabolites with inhibitory effects are ribosomally synthesized and post-translationally modified peptides (RiPPs), non-ribosomally synthesized peptides, type I polyketide synthase (PKS) products, type II PKS-derived polyketides, aminoglycosides and peptide-polyketide hybrids.

Our analysis of RiPP BGCs revealed a broad variety (103 BGCs) in the genomes of the *At*-LSPHERE strain collection (Fig. 3a).

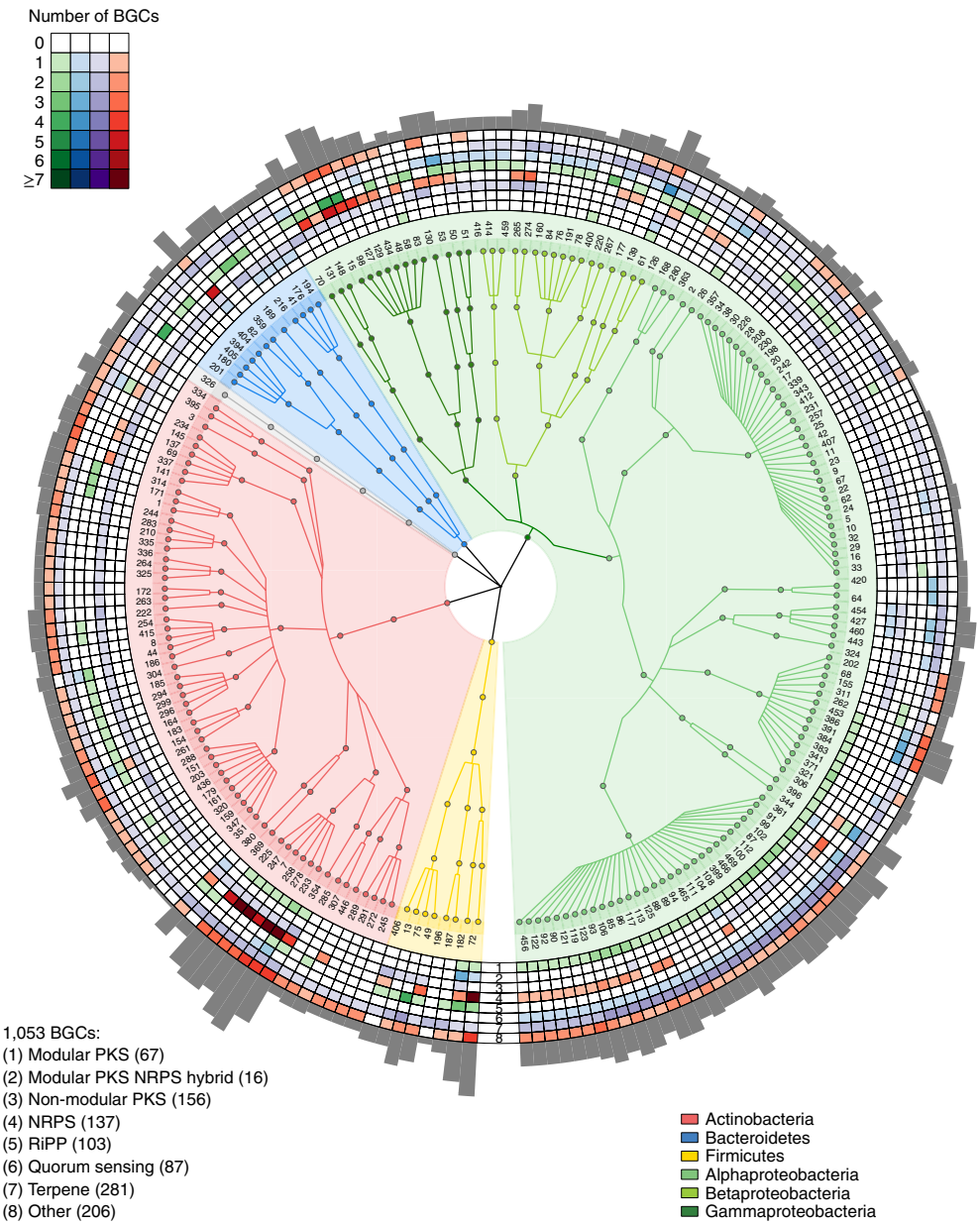

Fig. 2 | Biosynthetic potential of the *At*-LSPHERE strain collection. Taxonomic representation of the genome-sequenced *At*-LSPHERE strain collection ($n = 207$) with circular heat-map representation of number of detected BGCs and bar representation of total number of BGCs per isolate (Supplementary Table 9). Inner ring labels indicate strain numbers. Outer rings correspond to the following BGCs: (1) modular PKS, (2) modular PKS-NRPS hybrid, (3) non-modular PKS, (4) NRPS, (5) RiPP, (6) quorum sensing, (7) terpene, (8) other. Numbers in parentheses correspond to total number of BGCs in each category. Heat-map colours correspond to number of BGCs detected.

According to antiSMASH and BiG-SCAPE analyses, these belong to the RiPP families bacteriocins, lanthipeptides, lassopeptides, microviridins, linaridins, thiopeptides, thiopeptide-linaridin hybrids and lantipeptide-proteusin hybrids, indicating a broad structural diversity of ribosomally encoded peptides. Only a few lantipeptides and one bacteriocin show distant similarity to known RiPP BGCs. All other identified RiPP BGCs were almost exclusively family- or genus-specific and did not cluster with any characterized BGCs, thus indicating a large potential for the identification of different RiPP types. Many RiPPs are narrow-spectrum antibiotics[21] and hence might be responsible for selectively inhibiting only a few members of the strain collection.

antiSMASH and BiG-SCAPE analyses suggested the presence of mono-modular type I PKSs, particularly in members of the Rhizobiales and *Nocardiaceae* (Fig. 3b). Notably, in the genus *Rhodococcus*, a mono-modular type I PKS BGC is conserved across all strains of the collection, and a related BGC is present in *Willamsia*, another member of the *Nocardiaceae* (Supplementary Fig. 6a). All other type I PKSs and PKS-non-ribosomal peptide synthetase (NRPS) hybrids were only identified in individual genera, with strains of the phyla Firmicutes and Bacteroidetes being particularly rich in architecturally unusual multi-modular PKSs that suggest polyketides with unprecedented scaffolds (Fig. 3a, Supplementary Figs. 5a and 6b).

Trans-acyltransferase (*trans*-AT) PKS or hybrids of *trans*-AT PKS and NRPS BGCs were detected in the genomes of *Duganella* sp. Leaf61 and *Brevibacillus* sp. Leaf182 (Supplementary Figs. 5b and 6c). The latter seems to be a prolific producer of *trans*-AT

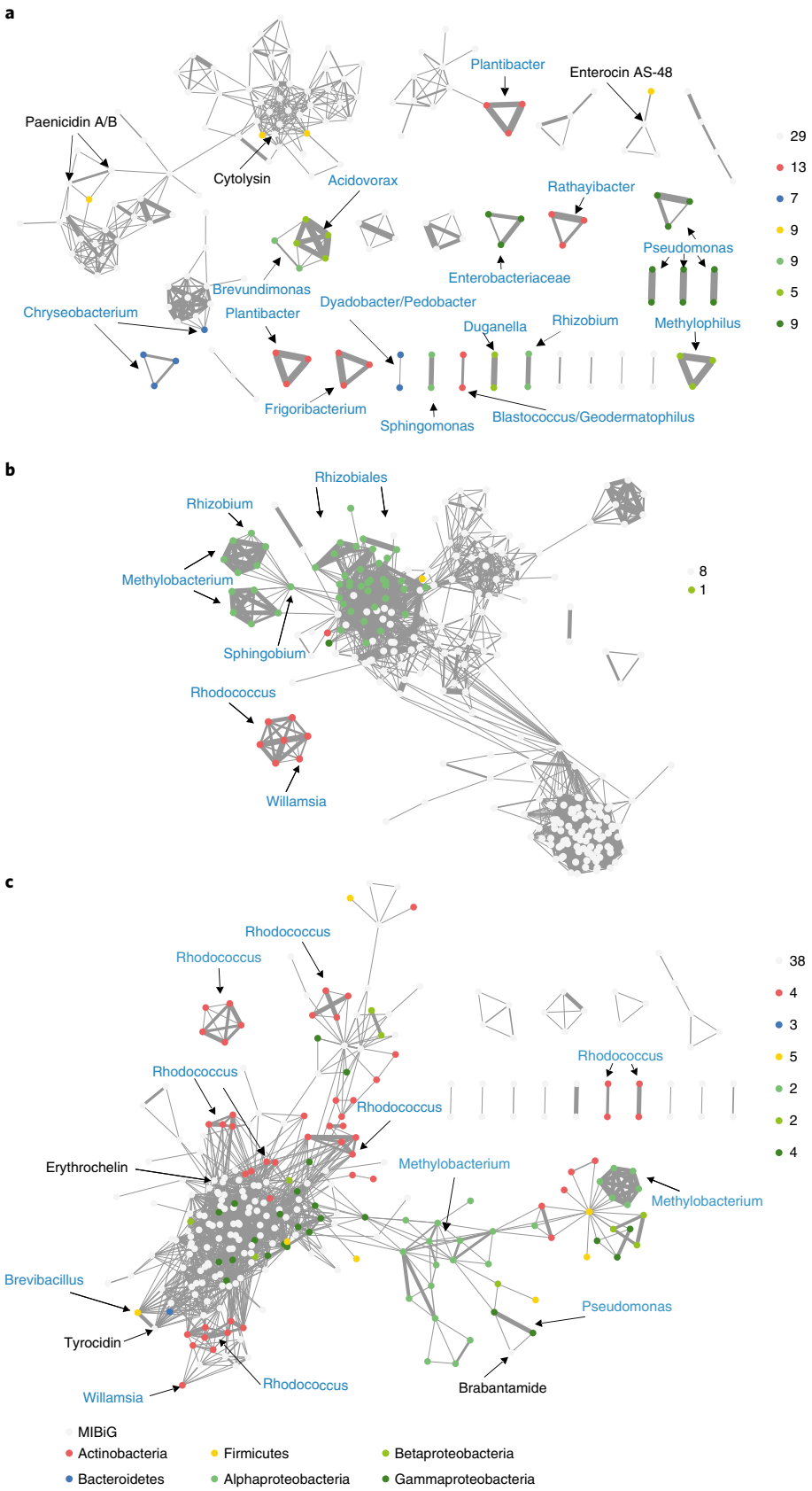

Fig. 3 | BiG-SCAPE analysis of BGCs detected by antiSMASH in 207 genomes of the *At*-LSPHERE strain collection and comparison with the MIBiG database of characterized BGCs. a, RiPP BGCs ($n=246$). **b**, Type I PKS BGCs ($n=277$). **c**, NRPS BGCs ($n=338$). Nodes and singletons are colour coded according to their origin (MIBiG database or phylum/class) and represent individual BGCs. The widths of interconnecting edges indicate the degree of relatedness between two BGCs, with connections up to a raw distance of 0.75 retained (Supplementary Table 12). Numbers of singletons are indicated. Black labels denote compounds associated with selected MIBiG BGCs. Blue labels highlight clusters of related BGCs or individual BGCs of interest by phylogenetic distribution.

PKS-derived polyketides, harbouring at least four *trans*-AT PKS gene-containing loci, one of the highest numbers for any known bacterium[22,23].

In addition, we identified one type II PKS BGC in *Bacillus* sp. Leaf406, which is rare since type II PKSs are almost exclusively reported from filamentous actinomycetes[24].

NRPS BGCs were abundant in genomes of the Beta- and Gammaproteobacteria (Fig. 2), as well as in the top inhibitors of the phyla Firmicutes and Bacteroidetes, which contain one or more NRPS or NRPS hybrid BGCs. Of all 135 NRPS and 18 NRPS hybrid BGCs, the majority (72 and 15, respectively) do not cluster with any MIBiG entry and thus suggest biosynthetic novelty (Supplementary Table 13, Fig. 3c). The family *Nocardiaceae* is noteworthy, being the only rich actinobacterial source of NRPS BGCs in the *At*-LSPHERE. A portion of the *Rhodococcus* NRPS BGCs are distantly related to characterized BGCs such as the erythrochelin biosynthetic locus[25], with which they share predicted adenylation domain substrate specificities. However, the majority of *Rhodococcus* NRPSs are (1) shared among members of the *At*-LSPHERE, (2) predicted to encode unusual peptides that are rich in Ser or Cys residues and (3) only show weak similarity to any MIBiG entry. Some of these shared BGCs encode up to 17 NRPS modules, of which 10 (60%) contain adenylation domains with predicted specificities for Ser or Cys (Supplementary Fig. 6d). One of these conserved BGCs shows distant similarity to the significantly smaller BGC for the Cys-rich cytotoxin thiocoralin[26], which in contrast encodes only five modules.

In addition, BiG-SCAPE analyses also revealed a putative mangotoxin BGC encoded in different *Pseudomonas* species (Supplementary Figs. 5c and 6e). This is noteworthy because mangotoxin is an antimetabolite that inhibits ornithine/arginine biosynthesis in plants and causes apical leaf necrosis[27].

We compared the total number of BGCs and the number of observed inhibitions for each strain (Supplementary Fig. 7). Most strains with potent inhibitory activities contain four or more BGCs of these potentially inhibitory BGC classes. An exception is *Exiguobacterium* sp. Leaf196 that is devoid of any PKS, NRPS or RiPP BGCs, yet caused 3 strong and 13 weak inhibitions, potentially based on either undetected BGCs, antibacterial proteins or competition for nutrients rather than antibiosis. Other examples include the *Nocardiacae* that based on our antiSMASH and Big-SCAPE analyses are particularly prolific NRPS producers, yet do not show any inhibitory activities, or *a Paenibacillus* sp. Leaf 72 that also harbours multiple BGCs classified into groups with putative antibiotic potential, yet no inhibitory activity was observed in our screen. This discrepancy might either be due to BGCs that are silent under the growth conditions used, or the products of the BGCs have a different ecological function beyond bacterial inhibition.

Besides BGCs that might be involved in inhibitory microbe–microbe interactions, clusters potentially involved in bacterial communication were detected (for example, homoserine lactone and butyrolactone BGCs[28]). Furthermore, we detected BGCs for compounds involved in environmental metal chelation (for example, NRPS-derived siderophores in addition to xanthoferrins, desferrioxamin and petrobactin-like siderophores) (Supplementary Figs. 5c and 6f)[29–32].

Terpenes are the most abundant BGC class, with at least 1 cluster detected in 187 (91%) strains (Fig. 2, Supplementary Fig. 5d) with the exception of Gammaproteobacteria that were almost completely devoid of detectable terpene BGCs. The high prevalence of BGCs for terpene biosynthesis partly relates to carotenoid and pigment biosynthesis. It has been postulated that pigmentation is favourable for leaf-colonizing bacteria, as they are exposed to solar radiation in their natural habitat[17,33,34]. For this reason, we examined whether genomes of strains isolated from leaves are enriched in pigment BGCs (carotenoid BGCs and arylpolyene PKSs[35]) compared to isolates obtained from the below-ground root and soil compartments of the *At*-SPHERE collection[16]. Candidates for pigment biosynthesis were detected in 181 leaf isolates (88% of collection; Supplementary Fig. 8, Supplementary Table 14) compared to 111 soil and root isolates (50% of collection), confirming that they are indeed significantly over-represented in genomes of leaf isolates (two-sided Fisher's exact test, $P < 2.2 \times 10^{-16}$). In addition to carotenoid BGCs, BiG-SCAPE analysis of leaf isolates revealed other terpene BGCs and a wealth of genus-specific terpene BGCs mainly in Actinobacteria and Proteobacteria (Supplementary Fig. 5d).

Generally, a number of BGCs (mainly terpene, type III PKS, mono-modular type I PKS and NRPS BGCs) are conserved amongst all members of a given genus, or even crossing genus boundaries. However, the top inhibitor strains contain BGCs that are unique to the producers. This distribution indicates that the corresponding metabolites could be responsible for the strong inhibitory potential observed for a particular strain.

Chemical analysis of the top inhibitor strains by matrix-assisted laser desorption/ionization (MALDI) imaging. The top inhibitors *Brevibacillus* sp. Leaf182, *Pseudomonas* sp. Leaf98, *Bacillus* sp. Leaf49 and *Flavobacterium* sp. Leaf82 were responsible for more than 30% of all inhibitory interactions. Each of these strains inhibited bacteria belonging to at least six different classes (Supplementary Table 3) and harboured a large number of BGCs putatively involved in the production of antibiotics (Supplementary Fig. 7). We thus selected these strains for chemical analysis using MALDI imaging mass spectrometry (IMS). Plate assays against Gram-positive and Gram-negative bacteria affirmed a multitude of colony-associated and secreted metabolites (Fig. 4a,b), supporting the notion that these strains are prolific producers of specialized metabolites. We analysed the IMS dataset for diffusable metabolites confined to the zone of inhibition (Fig. 4b, Supplementary Fig. 9). These analyses revealed the top inhibitor *Brevibacillus* sp. Leaf182 as the most conspicuous producer of candidate metabolites with various masses confined to the zone of inhibition. We therefore selected this strain for in-depth chemical analysis using a bioactivity-guided and IMS-assisted purification strategy.

Bioactivity-guided identification of antibiotics from *Brevibacillus* sp. Leaf182. To characterize the putative antibiotics produced by *Brevibacillus* sp. Leaf182, we fractionated liquid culture extracts and subsequently tested them for antibiotic activity against a panel of sensitive leaf strains. Four bioactive fractions were identified that contained metabolites belonging to the same compound family based on molecular network analysis[7,36] (Supplementary Fig. 10). A combination of tandem mass spectrometry and nuclear magnetic resonance (NMR) experiments followed by database search revealed streptocidin D (**1**) of the tyrocidin family as the bioactive principle in one of the fractions (Supplementary Figs. 11–16, Supplementary Table 15). Searching the genome for an NRPS BGC with matching adenylation domain substrate specificities resulted in the identification of an NRPS BGC with strong homology to the tyrocidin NRPSs (Supplementary Fig. 17, Supplementary Tables 12 and 16)[37]. In agreement with a previous study[38], streptocidin D (**1**) showed inhibition against most Gram-positive but not Gram-negative strains tested (Supplementary Table 17). Analysis of the mass spectrometry network data revealed at least nine congeners (Supplementary Figs. 10, 11 and 18–25). In addition to the known congeners streptocidins A (**2**), B/C (**3**), D (**1**), G (**4**) and M (**5**), we identified four previously undescribed variants that we termed streptocidins Q (**6**), R (**7**), S (**8**) and T (**9**) (Supplementary Figs. 18–25, Supplementary Table 18).

Because most of the metabolites identified in IMS experiments were not detected in liquid culture, we tested organic extracts from agar plates inoculated with *Brevibacillus* sp. Leaf182. Two different bioactive metabolites were purified in consecutive rounds of

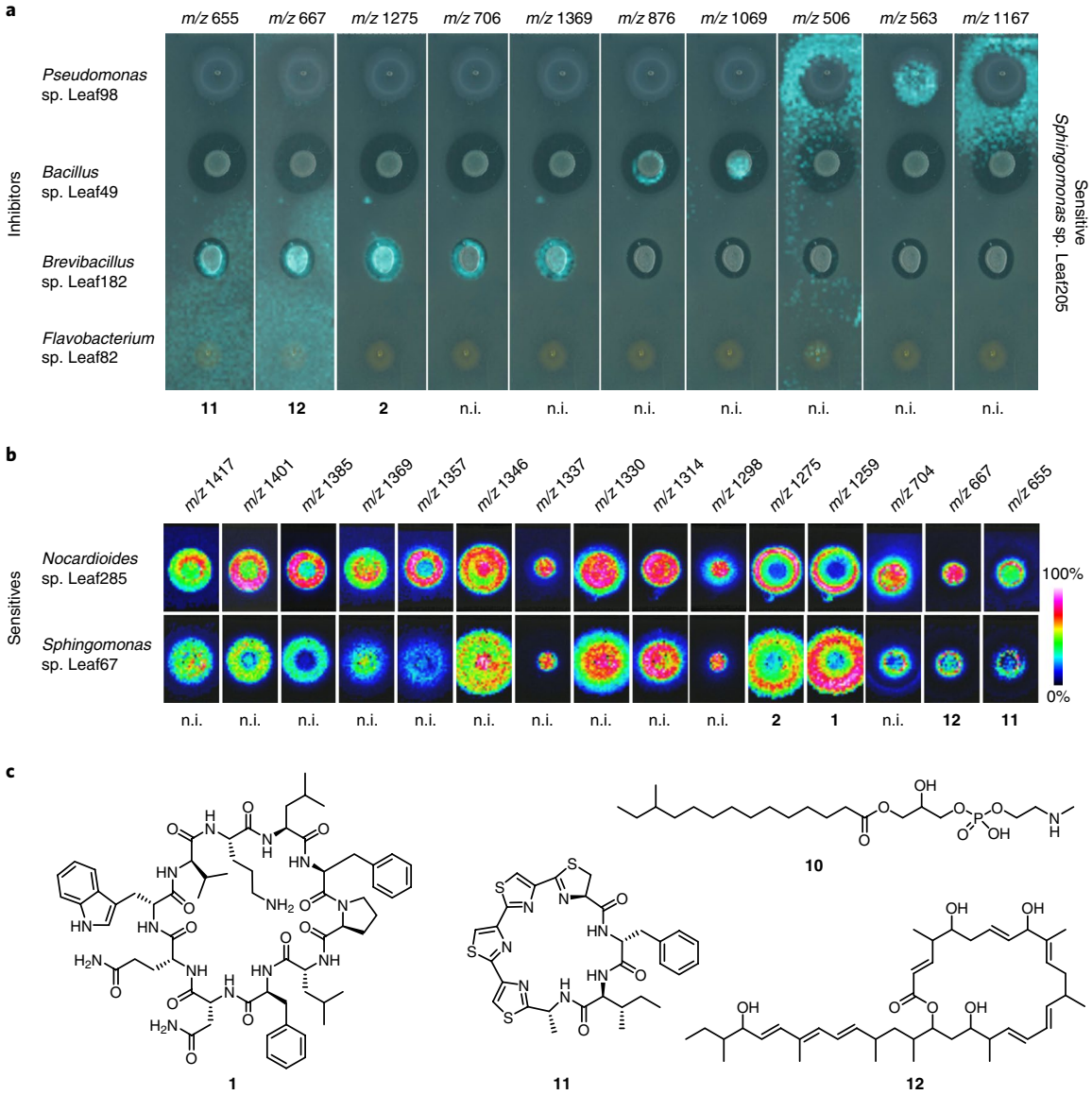

Fig. 4 | MALDI imaging results of selected top inhibitors against different sensitive strains and antibiotics isolated from *Brevibacillus* sp. Leaf182. a, False colour representation of selected metabolites (indicated by *m/z* values) produced by four of the top inhibitors against *Sphingomonas* sp. Leaf205 detected by MALDI imaging (one biological replicate). **b**, False-colour heatmap (low intensity, blue, to high intensity, red) representation of selected metabolites produced by the top inhibitor *Brevibacillus* sp. Leaf182 on *Nocardioides* sp. Leaf285 or *Sphingomonas* sp. Leaf67 confined to the zone of inhibition as detected by MALDI imaging (one biological replicate). **c**, Structures of bioactive metabolites isolated in this study. 1, streptocidin D; 2, streptocidin A; 10, phosphobrevin; 11, marthiapeptide A; 12, macrobrevin; n.i., not identified.

bioactivity-guided chromatographic separation, followed by mass spectrometry- and NMR-based structure elucidation. We identified an unusual lysophospholipid that we termed phosphobrevin (**10**) (Supplementary Figs. 26–31, Supplementary Table 19). The compound acts against Gram-negative bacteria (Supplementary Table 17), for which activities are particularly difficult to find[39]. The second antibiotic was identified as marthiapeptide A (**11**) (Supplementary Figs. 32–37, Supplementary Tables 17 and 20), which was previously isolated from the deep-sea bacterium *Marinactinospora thermotolerans*[40]. Despite genome sequencing efforts, the BGC for marthiapeptide A had remained enigmatic. We thus screened the genome of the reported producer and *Brevibacillus* sp. Leaf182 for a candidate BGC responsible for marthiapeptide biosynthesis. After genome resequencing of *Brevibacillus* sp. Leaf182 and de novo assembly, we identified an NRPS BGC as a candidate for marthiapeptide biosynthesis (Supplementary Fig. 38, Supplementary Table 21). To test whether the candidate BGC is indeed involved in marthiapeptide biosynthesis, the core biosynthetic gene *marB* was disrupted and the wild type and knockout subjected to comparative metabolomic analysis (Supplementary Fig. 39). These analyses confirmed that the *mar* BGC is responsible for marthiapeptide A biosynthesis.

Genome mining-guided identification of unusual bioactive metabolites. In addition to the activity-guided metabolite identification approach described above, we used a genome-mining strategy to exemplarily correlate genome to phenome. Again, we focused on *Brevibacillus* sp. Leaf182 due to its high number of loci encoding *trans*-AT PKSs, a family of natural product enzymes known for its potential to generate chemically diverse bioactive molecules[22].

Closer inspection revealed three of these are PKS-NRPS hybrid clusters and the fourth displays a *trans*-AT PKS and an NRPS in close vicinity. We concentrated on the latter *trans*-AT PKS BGC (named *bre* PKS; Supplementary Fig. 40, Supplementary Table 22) and subjected its protein sequences to the automated genome mining tool TransATor (transATor.ethz.ch/Trans-AT-PKS) (Supplementary Results 1). TransATor annotates *trans*-AT PKSs and predicts the structure of corresponding metabolites. The predicted structure (Supplementary Fig. 40b) was used to search for similar metabolites in natural product databases. Since no matching candidates were identified, the BGC was a strong candidate for the biosynthesis of a so far undescribed polyketide scaffold. *Brevibacillus* sp. Leaf182 cell extracts were used for mass spectrometry-guided purification, which was highly challenging as the compound was prone to degradation. Structure determination using mass spectrometry and two-dimensional NMR experiments identified a previously unknown polyketide that we named macrobrevin (**12**) (macro for macrolactone and brevin to indicate the producer *Brevibacillus* sp. Leaf182), displaying a unique chemical scaffold (Supplementary Figs. 41–47, Supplementary Table 23). Disruption of *breB* abolished macrobrevin production, as expected (Supplementary Fig. 48). We tested macrobrevin (**12**) against a panel of different *At*-LSPHERE isolates and identified the compound to be active against *Bacillus* sp. Leaf49 (Supplementary Table 17). Our genome mining-based polyketide discovery highlights the hidden metabolic diversity of the *At*-LSPHERE strain collection and suggests, together with the BGC data, that the phyllosphere harbours an untapped potential for previously unknown specialized metabolites, as exemplarily shown for *Brevibacillus* sp. Leaf 182.

Discussion

Nature is a prolific resource for specialized metabolites of pharmaceutical importance, such as antibiotics and anticancer drug leads[2]. In times of increased need for antimicrobial drugs, it is essential to move beyond traditionally screened habitats and microorganisms for the discovery of so far unknown natural product scaffolds[11]. For example, in an in silico and taxon-based approach, a recent study on members of the genus *Bacillus*, which also included various plant-derived strains, revealed diverse BGCs that suggested biosynthetic specialization[41]. Another comprehensive study revealed the presence of BGCs for thiopeptides, a common family of antibiotics, in human-associated bacteria[42]. Here, we applied a habitat- and interaction-centred strategy to evaluate the potential for unprecedented antimicrobials. Microbial inhabitants of the phyllosphere have to compete for the scarcely available nutrients and trace elements and are exposed to ultraviolet radiation, oxidative stress, drought and fast-changing conditions[17]. The adaptation to this environment might be a driving force for the selection of microorganisms that produce potent specialized metabolites to compete with other bacteria and cope with harsh environmental conditions. To date, the leaf environment has not been systematically analysed in this regard. We combined a high-throughput binary interaction screen of a representative strain collection of more than 200 leaf isolates (*At*-LSPHERE[16]) and genome mining. We identified more than 1,000 BGCs for compounds putatively involved in bacterial interactions and niche adaptation, with a majority lacking similarity to characterized BGCs. These belong to diverse biosynthetic classes including RiPP and terpene systems, NPRSs and *trans*-AT PKSs, suggesting high potential for uncharacterized natural products. This hypothesis is supported by isolating two unknown antibiotics, the lysolipid phosphobrevin and the *trans*-AT PKS-derived macrobrevin, from one selected strain of the *At*-LSPHERE collection.

We observed that a few strains of the *At*-LSPHERE collection showed a wide inhibition spectrum, which is consistent with two binary interaction screens using 67 bacterial isolates from 8 environments[43] and 140 bacterial isolates from different compartments of *Echinaceae purpurae*[44], respectively. The total number of antagonistic interactions observed ranged from 1.4% of all combinations (this study) to 3.4%[43] and 11.3%[44]. However, a direct comparison should be taken with caution because a different method to examine inhibitions was used in the latter study that consisted of cross-streaking (might also detect inhibition principles other than natural products, such as modification of pH or nutrient depletion) versus direct plating with detection of inhibition halos.

We explored unknown compounds in an exemplary phyllosphere habitat. In addition to implications for natural product discovery, the data provide a valuable basis for testing the relevance of natural products for community assembly. It is assumed that antimicrobial compounds contribute to the plant microbiota composition[45,46]. Our collection of cultivated, genome-sequenced bacteria will permit in planta studies to test the impact of individual BGCs in wild-type and mutant strains within complex synthetic communities[15]. The *Brevibacillus* mutants generated here represent a good starting point to study how specialized metabolism shapes plant microbiomes[47]. Natural products synthesized by the plant microbiota might act on the producer or other microbes and/or the plant[48]. In addition, natural product effects might be dose-dependent in situ[49]. In this context, in planta studies will also be informative with regard to BGC-rich strains that did not show an antagonistic interaction.

In conclusion, we have shown that the phyllosphere is a promising resource for bacteria with a large and distinct biosynthetic repertoire that provides the basis for the isolation of bioactive metabolites with unusual structural scaffolds and in planta testing. Overall, a combined strategy of ecosystem-, activity- and genome-guided analyses as conducted here is a promising path for natural product research in the context of ecological studies and compound discovery.

Methods

Pair-wise interaction screen of phyllosphere bacteria. The 224 leaf isolates of the *At*-LSPHERE[16] were grown on R2A agar (R-2A, Sigma Aldrich) supplemented with 0.5% (v/v) methanol (R2A + M) or minimal medium[50] containing 0.5% (v/v) methanol and 25 mM glucose (MM + M + G) as carbon sources at a pH of 6.5 at room temperature (for plates) or 28 °C (for liquid cultures) (Supplementary Table 24). Strains tested for sensitivity were restreaked from cryostocks, incubated on solid media and used for the inoculation of 5 ml liquid cultures in round-bottom cultivation tubes (Falcon). High-throughput cultivations of inhibitory isolates were performed in 96-well plates containing solid media. To assess binary interactions between the different strains, stationary liquid cultures of strains tested for sensitivity were diluted to a total of 50 ml with melted R2A + M agar at 45 °C (final agar concentration ≈ 1.35% (w/v)) and 25 ml of each were poured into 2 square plates to form a uniform layer containing the test strain. For strains that did not grow sufficiently well in either liquid medium or when poured into R2A + M agar, the sensitives were grown on R2A + M agar plates and resuspended in 5 ml liquid R2A + M before pouring the lawn. The inhibitory interaction partners were resuspended from 96-well plates and incubated for 2 h while shaking. Subsequently, roughly 1 µl of each strain on the 96-well plate was printed onto the solidified agar layer containing the putative sensitive strain using a replicator. Results of the binary interactions were evaluated after 2–3 d incubation at room temperature. Interactions were classified as 'strong inhibitions' if the inhibition zone, the distance of the outside border of the halo to the edge of the colony, exceeded 3 mm, while 'weak inhibitions' either showed a smaller or partially turbid halo (Fig. 1a). Data analysis, statistics and visualization were done using R, Python (including the NetworkX package) and GraPhlAn[51].

Pair-wise interaction screen of selected phyllosphere bacteria on different media. A subset of ten indicator strains (see Supplementary Table 25) was tested for sensitivities against the *At*-LSPHERE collection on six different media. Briefly, these were a barley extract medium (BEM), a plant-mimic medium (PMM) at two different pHs, half-strength lysogeny broth (½ LB) supplemented with 0.5% (*v*/*v*) methanol and R2A with and without methanol supplementation (for details on media composition see Supplementary Methods). These media were chosen to mimic different growth conditions. BEM and PMM were designed to mimic the plant surface, the first being a full plant extract and thus containing all of its nutrients only supplemented with methanol and the latter a synthetic medium containing sugars, small organic acids and amino acids probably encountered on plant surfaces[18,52] as well as glycerol. The ½ LB is richer in nutrients and contains more salt compared to the standard medium R2A + M. Strains grown on R2A + M agar were suspended in 10 mM $MgCl_2$ solution and inoculated onto each medium

to be tested in 96-well plates ('inhibitior strains', complete *At*-LSPHERE collection) or on round agar plates ('sensitive strains', see Supplementary Table 25). Strains were grown at 22 °C for 4 d (BEM, R2A and R2A + M) or 5 d (PMM 6.1, PMM 7.1 and ½ LB). Sensitive strains were resuspended in each medium and an amount corresponding to a final optical density $(OD)_{600}$ of 0.01 was added to 15 ml top agar (42 °C, 1% agar) and immediately poured on top of a square plate containing 25 ml of the same medium (1.5% agar). Inhibitor strains were resuspended by adding liquid medium to each well of the 96-well plate, incubating for 15 min and then vortexing for 5 min. Bacterial suspensions were transferred to empty 96-well plates and printed onto the solidified agar layer containing the putative sensitive strain using a replicator as described above. Appearance of inhibitory halos was scored after incubation at 22 °C for 2–3 d as described above. Interactions were not scored when either one of the interaction partners had not grown or when the strain spotted on top showed a contamination.

Genome sequencing of *Pseudomonas* sp. Leaf98 and *Brevibacillus* sp. Leaf182. Genomic DNA of *Pseudomonas* sp. Leaf98 was extracted using the MasterPure DNA Purification kit (Epicentre). For *Brevibacillus* sp. Leaf182, genomic DNA was extracted as described in reference[53]. Libraries were prepared from purified genomic DNA and sequenced using the Illumina HiSeq platform. Draft genomes were de novo assembled using SPAdes 3.11 for *Pseudomonas* sp. Leaf98 and CLC Genomics Workbench 10 for *Brevibacillus* sp. Leaf182 (Supplementary Methods).

Secondary metabolite cluster prediction. We employed the antiSMASH standalone toolkit v.4.0.2[19] to mine the genomes of all genome-sequenced leaf isolates (BioProject PRJNA297956, QFZI01000000, LMPN02000000) for the presence of putative natural product BGCs. The identified BGCs were grouped (Supplementary Table 10), summarized for each strain (Supplementary Table 9) and visualized using GraPhlAn[51].

BiG-SCAPE analysis. BGCs from antiSMASH analyses were compared to the MIBiG database[20] v1.3 using BiG-SCAPE (BiG-SCAPE-master-596cfbe250563053 79e0a05e8442492891197c32; downloaded 19 February 2018) with PFAM database 31.0[54]. Analysis was conducted using default settings with mode 'auto' and retaining singletons. Networks were computed for raw distance cut-offs of 0.10–0.80 in increments of 0.05. The lower the cut-off, the fewer connections are kept between clusters[55]. Results (Supplementary Table 12) were visualized as a network using Cytoscape 3.6 for a cut-off of 0.75 (Fig. 3, Supplementary Fig. 5).

TransATor-based analysis of *trans*-AT PKS BGCs. Protein sequences of *trans*-AT PKSs identified by antiSMASH were extracted and subjected to the automated *trans*-AT PKS genome mining and structure prediction pipeline TransATor (transATor.ethz.ch/Trans-AT-PKS). The structures predicted by TransATor were used for similarity searches in MarineLit (http://pubs.rsc.org/marinlit/), ChemSpider (http://www.chemspider.com/) and the Dictionary of Natural Products (http://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml).

Analysis of carotenoid and arylpolyene BGCs. Arylpolyene BGCs were identified by antiSMASH analysis (v.4.0.2)[19] for the genomes available from the *At*-SPHERE collection[16] (BioProjects PRJNA297956, PRJNA297942, PRJNA298127; QFZI01000000). Strains putatively containing carotenoid biosynthesis genes were predicted based on the presence of homologues of at least two of the three proteins: phytoene synthase, phytoene desaturase and lycopene cyclase. To identify potential phytoene synthase and lycopene cyclase proteins, we queried the *At*-SPHERE strain collection database using the HMMER toolkit (http://hmmer.org/, v3.1b2) with the hidden Markov models PF00494.18 and PF05834.1, respectively. Proteins with an e-value threshold of 1×10^{-20} and 1×10^{-25} were considered similar for the phytoene synthase and lycopene cyclase, respectively. To identify potential phytoene desaturase proteins, we used blastp of the BLAST+ standalone software (v.2.2.31) to query the sequence ACS41376.1 of *Methylobacterium extorquens* AM1 against the *At*-SPHERE database. Proteins with an *e*-value threshold of 1×10^{-25} were considered a hit. Fisher's exact test was used to compare the presence of either an arylpolyene BGC or carotenoid biosynthesis genes in the *At*-LSPHERE and the remaining *At*-SPHERE.

IMS of bacterial colonies. Sensitive indicator strains were grown on R2A + M, resuspended in 10 mM $MgCl_2$, diluted into 25 ml melted R-2A + M top agar (1% agar, 42 °C) to a final OD_{600} of 0.01 and immediately poured into a petridish (diameter 13.5 cm). Inhibitory strains were resuspended in 10 mM $MgCl_2$ at an OD_{600} of 4 and 2 µl each spotted on top. After incubation at room temperature for 2 d, samples were prepared and imaging experiments conducted as described (for details see Supplementary Methods)[56].

Isolation, purification and structure elucidation of streptocidin D and macrobrevin from liquid culture extracts. *Brevibacillus* sp. Leaf182 was cultured in liquid R2A + M medium in 1 l ultra-high-yield flasks (Thomson Instrument Company) for 2–4 d at 28 °C. *Brevibacillus* sp. Leaf182 culture (3 l) was harvested, and the pellet was extracted with acetone and the resulting extract subjected to reversed-phase high-performance liquid chromatography (RP-HPLC) (Phenomenex Kinetex 5 µm C18 (100 A, 250 × 10 mm) column, ultraviolet detection at λ = 280 nm, room temperature, with 5% MeCN for 5 min, then a gradient from 5% MeCN to 100% MeCN for 30 min and 100% MeCN for 25 min).

For the isolation of streptocidin D, inhibitory fractions 9–13 (eluting at 24–36 min) were subjected to ultra-high-performance liquid chromatography data-dependent mass spectrometry[2] analysis coupled to molecular network analysis as previously described (for details see Supplementary Methods)[36,57]. For the isolation of streptocidin D, fraction 11 (eluting at 30–33 min) was fractionated by RP-HPLC using a Phenomenex Luna 5 µm C18 (100 A, 250 × 21.2 mm) column on an Infinity 1260 instrument (Agilent, Santa Clara CA, USA) with a water to acetonitrile gradient of 0–25% in 10 mins, 25–65% in 36 mins and 65–100% in 10 mins.

For the isolation of macrobrevin, the fraction containing the metabolite matching the structure predicted by TransATor (see above, fraction eluting at 46-48 min) was further purified by RP-HPLC (Phenomenex Luna 5 µm Phenyl-Hexyl (100 A, 250 × 10 mm) column, ultraviolet detection at λ = 280 nm, room temperature, 70% MeCN) to yield 0.5 mg macrobrevin. For details on mass spectrometry- and NMR-assisted structure elucidation see Supplementary Methods.

Isolation, purification and structure elucidation of marthiapeptide A and phosphobrevin from agar plate extractions. *Marthiapeptide A:. Brevibacillus* sp. Leaf182 was cultured on R2A + M agar (50 square petri dishes) and incubated for 2 d at room temperature. Agar plates were extracted with ethyl acetate. The extracts were filtered and dried under reduced pressure. Crude extracts were redissolved in a 2 ml methanol-water solution (1:1), centrifuged and the pellet discarded. Inhibitory activities of supernatants from *Brevibacillus* sp. Leaf182 agar plate extracts were determined by filter disc assays against a panel of leaf isolates. Briefly, sterilized filter discs (Macherey-Nagel, MN615) were loaded with 7 µl extracts, dried and placed onto R2A + M agar containing a sensitive strain as described above. Active extracts were fractioned by RP-HPLC using a Phenomenex Luna 5 µm C18 (100 A, 250 × 21.2 mm) column, 0–60% in 10 mins, 60–100% in 20 mins, 100% for 5 mins and a flow rate of 21.2 ml min^{-1}, monitored by ultraviolet absorption. The collected fractions were tested for inhibitory activity as above. The second round of purification was conducted using a Phenomenex Luna 5 µm Phenyl-Hexyl (100 A, 250 × 21.2 mm) column. The gradient was set to 0–60% in 10 mins and 60–100% in 36 mins.

Phosphobrevin:. Individual colonies of *Brevibacillus* sp. Leaf182 (5 µl each) were spotted on R2A + M agar (170 petri dishes with approximately 40 colonies each) and incubated for 4 d at room temperature. Agar plates were lyophilized and extracted with 6 l of a methanol/water mixture (80:20). The extracts were filtered and dried under reduced pressure. Extracts were redissolved in water and consecutively extracted with hexane, chloroform, diethyl ether and ethyl acetate. Inhibitory activities were determined by filter disc assays against a panel of leaf isolates. A quarter of the active ethyl acetate fraction was fractionated by preparative RP-HPLC using a Phenomenex Kinetex 5 µm XB-C18 (100 A, 150 mm × 4.6 mm) column with a water to acetonitrile (ACN) gradient (10–100% ACN in 35 mins, 100% ACN for 7 mins and a flow rate of 21.2 ml min^{-1}) monitored by ultraviolet absorption. Fractions were collected every minute and tested for inhibitory activities as above. The bioactive fraction 24 was subjected to a second round of reverse-phase chromatography-based separation using a Luna 5 µm Phenyl-Hexyl (100 A, 250 × 21.2 mm) column with a water to acetonitrile gradient. The gradient was set to 25% ACN for 5 mins, 25–100% ACN in 45 mins and 100% ACN for 5 mins and a flow rate of 21.2 ml min^{-1} monitored by ultraviolet absorption. The bioactive fraction 24 was subjected to mass spectrometry- and NMR-assisted structure elucidation (Supplementary Methods).

Confirmation of marthiapeptide A and macrobrevin BGCs in *Brevibacillus* sp. Leaf182 by marker exchange mutagenesis. Homologous regions up- and downstream of the genes of interest (*breB* nucleotides 1–1839, *marB* nucleotides 1–2699) were amplified from *Brevibacillus* sp. Leaf182 genomic DNA and introduced into a pSEVA281[58] vector using the oligonucleotides and restriction sites indicated in Supplementary Table 26. The erythromycin resistance cassette (*erm*) was amplified from pBs2E[59] and inserted into pSEVA281_hr1/2 yielding the final knockout plasmids. Non-replicative knockout plasmids were transferred into *Brevibacillus* sp. Leaf182 by conjugation with *Escherichia coli* S17 (λ-pir) and transformants were selected on ½ LB supplemented with colistin (10 µg ml^{-1}) and erythromycin (5 µg ml^{-1}). Double crossover mutants were identified by PCR.

For confirmation of absence of macrobrevin and marthiapeptide, *Brevibacillus* sp. Leaf182 wild type and mutants (*Brevibacillus* sp. Leaf182 *breB*::*erm* and *Brevibacillus* sp Leaf182 *marB*::*erm*) were treated as described above for the isolation of macrobrevin and marthiapeptide, respectively, and the extracts subjected to ultra high performance liquid chromatography-high resolution-heated electrospray ionization-mass spectrometry experiments using the liquid chromatography solvent gradients described in the respective section for the

initial detection of the compounds on a Phenomenex Kinetex 2.6 µm C18 (100 A, 150 × 4.6 mm) column at a flow rate of 0.5 ml min^{-1} connected to a QExactive (Thermo scientific) mass spectrometer.

NMR. NMR spectra were recorded on a Bruker Avance III spectrometer equipped with a cold probe at 600 MHz for ^{1}H NMR and 150 MHz for ^{13}C NMR. Chemical shifts were referenced to the solvent peak at δ_H 7.27 and δ_C 77.23 for chloroform-*d*, δ_H 2.5 δ_C 39.51 for DMSO-d_6, δ_H 1.94 δ_C 1.39 for acetonitrile-d_3 and δ_H 3.31, δ_C 49.15 for methanol-d_4.

Bioactivity assay with purified compounds. Bacteria grown on R2A + M were resuspended in 10 mM $MgCl_2$, diluted into melted R2A + M top agar (1% agar, 42 °C) to a final OD_{600} of 0.01 and poured immediately on top of previously prepared R2A + M agar plates. The purified compounds were dissolved in MeOH (streptocidin D, phosphobrevin), 90% MeCN (marthiapeptide A) or MeCN (macrobrevin). Solvent control and purified compound (10 µl each) were applied on sterile filter paper discs (Macherey-Nagel, MN615), dried for 5–10 min under laminar flow and subsequently placed on the top agar plates containing bacteria. Plates were incubated at room temperature and formation of zones of inhibition was documented for 3–5 d (Supplementary Table 17).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. Source code of the TransATor web application is available from the corresponding authors upon request.

Data availability. Source data for Figs. 1, 2, 3 and 4 and Table 1 are available in the Supplementary Information files. Data supporting the findings of this study are generally available within the paper and its Supplementary Information files. The whole genome shotgun projects of *Pseudomonas* sp. Leaf98 and *Brevibacillus* sp. Leaf182 have been deposted at DDBJ/ENA/Genbank under the accessions QFZI00000000 and LMPN00000000. The versions described in this paper are QFZI01000000 and LMPN02000000. Raw reads have been deposited at SRA (SRP148446, SRP149408). Information on the marthiapeptide (BGC0001469) and macrobrevin (BGC0001470) BGCs was uploaded to the MIBiG database. All other raw data are available from the corresponding authors upon request.

Received: 22 September 2017; Accepted: 18 June 2018;
Published online: 23 July 2018

References

1. Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach* 3rd edn (Wiley, Chichester, 2009).
2. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661 (2016).
3. Mlot, C. Antibiotics in nature: beyond biological warfare. *Science* **324**, 1637–1639 (2009).
4. Meiser, P., Bode, H. B. & Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl Acad. Sci. USA* **103**, 19128–19133 (2006).
5. Hawver, L. A., Jung, S. A. & Ng, W. L. Specificity and complexity in bacterial quorum-sensing systems. *FEMS Microbiol. Rev.* **40**, 738–752 (2016).
6. Höfer, I. et al. Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chem. Biol.* **18**, 381–391 (2011).
7. Phelan, V. V., Liu, W. T., Pogliano, K. & Dorrestein, P. C. Microbial metabolic exchange—the chemotype-to-phenotype link. *Nat. Chem. Biol.* **8**, 26–35 (2012).
8. Saha, R., Saha, N., Donofrio, R. S. & Bestervelt, L. L. Microbial siderophores: a mini review. *J. Basic Microbiol.* **53**, 303–317 (2013).
9. Wilson, M. C. et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58–62 (2014).
10. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem.* **49**, 2011–2013 (2010).
11. Pidot, S. J., Coyne, S., Kloss, F. & Hertweck, C. Antibiotics from neglected bacterial sources. *Int. J. Med. Microbiol.* **304**, 14–22 (2014).
12. Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem. Biol.* **20**, 636–647 (2013).
13. Rondon, M. R. et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
14. Banik, J. J. & Brady, S. F. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* **13**, 603–609 (2010).
15. Vorholt, J. A., Vogel, C., Carlström, C. I. & Müller, D. B. Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* **22**, 142–155 (2017).
16. Bai, Y. et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**, 364–369 (2015).
17. Vorholt, J. A. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **10**, 828–840 (2012).
18. Ryffel, F. et al. Metabolic footprint of epiphytic bacteria on *Arabidopsis thaliana* leaves. *ISME J.* **10**, 632–643 (2016).
19. Blin, K. et al. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **45**, W36–W41 (2017).
20. Medema, M. H. et al. Minimum information about a biosynthetic gene cluster. *Nat. Chem. Biol.* **11**, 625–631 (2015).
21. Yang, S. C., Lin, C. H., Sung, C. T. & Fang, J. Y. Antibacterial activities of bacteriocins: application in foods and pharmaceuticals. *Front. Microbiol.* **5**, 241 (2014).
22. Helfrich, E. J. & Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
23. Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y. & Weber, T. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* **45**, D555–D559 (2017).
24. Hillenmeyer, M. E., Vandova, G. A., Berlew, E. E. & Charkoudian, L. K. Evolution of chemical diversity by coordinated gene swaps in type II polyketide gene clusters. *Proc. Natl Acad. Sci. USA* **112**, 13952–13957 (2015).
25. Lazos, O. et al. Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters. *Chem. Biol.* **17**, 160–173 (2010).
26. Lombó, F. et al. Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *Streptomyces* species. *ChemBioChem* **7**, 366–376 (2006).
27. Arrebola, E. et al. Mangotoxin: a novel antimetabolite toxin produced by *Pseudomonas syringae* inhibiting ornithine/arginine biosynthesis. *Physiol. Mol. Plant Pathol.* **63**, 117–127 (2003).
28. Bassler, B. L. & Losick, R. Bacterially speaking. *Cell* **125**, 237–246 (2006).
29. Pandey, S. S., Patnana, P. K., Rai, R. & Chatterjee, S. Xanthoferrin, the α-hydroxycarboxylate-type siderophore of *Xanthomonas campestris* pv. *campestris*, is required for optimum virulence and growth inside cabbage. *Mol. Plant Pathol.* **18**, 949–962 (2017).
30. Barona-Gomez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J. & Challis, G. L. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J. Am. Chem. Soc.* **126**, 16282–16283 (2004).
31. Lee, J. Y. et al. Biosynthetic analysis of the petrobactin siderophore pathway from *Bacillus anthracis*. *J. Bacteriol.* **189**, 1698–1710 (2007).
32. Barry, S. M. & Challis, G. L. Recent advances in siderophore biosynthesis. *Curr. Opin. Chem. Biol.* **13**, 205–215 (2009).
33. Lindow, S. E. & Brandl, M. T. Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* **69**, 1875–1883 (2003).
34. Lindow, S. E. & Leveau, J. H. Phyllosphere microbiology. *Curr. Opin. Biotechnol.* **13**, 238–243 (2002).
35. Schöner, T. A. et al. Aryl polyenes, a highly abundant class of bacterial natural products, are functionally related to antioxidative carotenoids. *ChemBioChem* **17**, 247–253 (2016).
36. Wang, M. et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **34**, 828–837 (2016).
37. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997).
38. Gebhardt, K., Pukall, R. & Fiedler, H. P. Streptocidins A-D, novel cyclic decapeptide antibiotics produced by *Streptomyces* sp. Tu 6071. I. Taxonomy, fermentation, isolation and biological activities. *J. Antibiot.* **54**, 428–433 (2001).
39. Zgurskaya, H. I., Löpez, C. A. & Gnanakaran, S. Permeability barrier of Gram-negative cell envelopes and approaches to bypass it. *ACS Infect. Dis.* **1**, 512–522 (2015).
40. Zhou, X. et al. Marthiapeptide A, an anti-infective and cytotoxic polythiazole cyclopeptide from a 60 L scale fermentation of the deep sea-derived *Marinactinospora thermotolerans* SCSIO 00652. *J. Nat. Prod.* **75**, 2251–2255 (2012).
41. Grubbs, K. J. et al. Large-scale bioinformatics analysis of *Bacillus* genomes uncovers conserved roles of natural products in bacterial physiology. *mSystems* **2**, e00040-17 (2017).
42. Donia, M. S. et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).
43. Russel, J., Røder, H. L., Madsen, J. S., Burmølle, M. & Sørensen, S. J. Antagonism correlates with metabolic similarity in diverse bacteria. *Proc. Natl Acad. Sci. USA* **114**, 10684–10688 (2017).

44. Maida, I. et al. Antagonistic interactions between endophytic cultivable bacterial communities isolated from the medicinal plant *Echinacea purpurea*. *Environ. Microbiol.* **18**, 2357–2365 (2016).
45. Hassani, M. A., Durán, P. & Hacquard, S. Microbial interactions within the plant holobiont. *Microbiome* **6**, 58 (2018).
46. Venturi, V. & Keel, C. Signaling in the rhizosphere. *Trends Plant Sci.* **21**, 187–198 (2016).
47. Chodkowski, J. L. & Shade, A. A synthetic community system for probing microbial interactions driven by exometabolites. *mSystems* **2**, e00129-17 (2017).
48. Stringlis, I. A., Zhang, H., Pieterse, C. M. J., Bolton, M. D. & de Jonge, R. Microbial small molecules—weapons of plant subversion. *Nat. Prod. Rep.* **35**, 410–433 (2018).
49. Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu. Rev. Phytopathol.* **50**, 403–424 (2012).
50. Peyraud, R. et al. Demonstration of the ethylmalonyl-CoA pathway by using ^{13}C metabolomics. *Proc. Natl Acad. Sci. USA* **106**, 4846–4851 (2009).
51. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029 (2015).
52. Müller, D. B., Schubert, O. T., Rost, H., Aebersold, R. & Vorholt, J. A. Systems-level proteomics of two ubiquitous leaf commensals reveals complementary adaptive traits for phyllosphere colonization. *Mol. Cell. Proteom.* **15**, 3256–3269 (2016).
53. Yamanaka, K. et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl Acad. Sci. USA* **111**, 1957–1962 (2014).
54. Finn, R. D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, 279–285 (2016).
55. Ceniceros, A., Dijkhuizen, L., Petrusma, M. & Medema, M. H. Genome-based exploration of the specialized metabolic capacities of the genus *Rhodococcus*. *BMC Genom.* **18**, 593 (2017).
56. Yang, J. Y. et al. Primer on agar-based microbial imaging mass spectrometry. *J. Bacteriol.* **194**, 6023–6028 (2012).
57. Ueoka, R. et al. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat. Chem. Biol.* **11**, 705–712 (2015).
58. Durante-Rodríguez, G., de Lorenzo, V. & Martínez-García, E. The Standard European Vector Architecture (SEVA) plasmid toolkit. *Methods Mol. Biol.* **1149**, 469–478 (2014).
59. Radeck, J. et al. The *Bacillus* BioBrick Box: generation and evaluation of essential genetic building blocks for standardized work with *Bacillus subtilis*. *J. Biol. Eng.* **7**, 29 (2013).

Acknowledgements

This work was financially supported by SNF grant NRP72 to J.P. and J.A.V. and by European Research Council Advanced Grants (PhyMo to J.A.V. and SynPlex to J.P.).

Author contributions

E.J.N.H., C.M.V., R.U., M.S., F.R., D.B.M., J.P. and J.A.V. designed the research. C.M.V., M.S., F.R., D.B.M. and M.K. performed binary interaction screens. E.J.N.H., C.M.V., F.R. and S.P. performed genome mining studies. C.M.V. and D.B.M. conducted statistical analyses. E.J.N.H., C.M.V. and M.S. conducted MALDI imaging experiments. E.J.N.H., C.M.V., M.S., F.R. and S.P. conducted bioassays. E.J.N.H., C.M.V., R.U., F.R. and S.P. isolated and structure-elucidated metabolites. M.S. generated *Brevibacillus* knockout mutants. E.J.N.H., C.M.V., D.B.M., J.P. and J.A.V. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-018-0200-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.P. or J.A.V.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Julia A. Vorholt; Jörn Piel

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample size was based on the published At-SPHERE strain collection (Bai et al. 2015 Nature).

2. Data exclusions

Describe any data exclusions.

We did not score interactions as "strong"/"weak"/"none" when either one of the interaction partners had not grown under the conditions used or when the strain spotted on top showed a contamination (applies to pairwise interaction screen for selected phyllosphere bacteria on different media).

3. Replication

Describe whether the experimental findings were reliably reproduced.

n/a

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

n/a

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

n/a

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Genome data was analyzed using antiSMASH, BiG-SCAPE, TransATor, BLAST+ and HMMER3. Emzed2 (Python), R and Excel were used for data analysis. The MiBIG database was used as reference. MS data was analyzed using Xcalibur, MALDI imaging data using FlexImaging and FlexAnalysis. NMR data was analyzed using TopSpin. Data was plotted using Cytoscape, R, Graphlan, NetworkX, Graphpad PRISM and MatPlotLib. Figures were assembled and finalized in Adobe Illustrator, Adobe Photoshop and PowerPoint.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	Purified compounds are not available for distribution. At-LSPHERE strains (Bai et al. 2015) are available from the DSMZ; mutant strains from J. Vorholt.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	n/a

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	n/a
b. Describe the method of cell line authentication used.	n/a
c. Report whether the cell lines were tested for mycoplasma contamination.	n/a
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	n/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	n/a

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	n/a